

Thoughts on Version 7

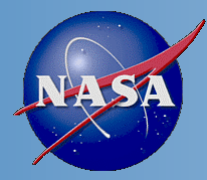
NASA Sounder Team Meeting

(NOTE: This presentation draws on some conclusions shown in the previous presentation (CrIMSS EDR status))

Christopher Barnet
NOAA/NESDIS/STAR

Nov. 15, 2012

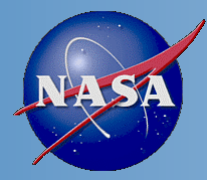




Objective



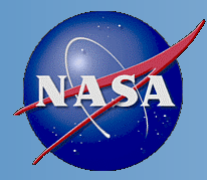
- This is a philosophical presentation intended to incite discussion on potential version 7 systems.
 - Primary concern is that users may not be aware of subtle characteristics of our products.
 - Primary goals are (1) encourage community acceptance of the AIRS products and (2) further exploit AIRS information content.
 - It is also possible that v7 could contain multiple product types (*e.g.*, one for climate and one for weather applications).
- My opinion of product attributes is not intended to offend any algorithm developer
 - Although, maybe it is more accurate to say I am trying to offend all algorithm developers equally.
- This discussion is at a high level (*i.e.*, no equations)
 - But, obviously, a primary objective of this talk is to discuss options in a mathematically rigorous manner.



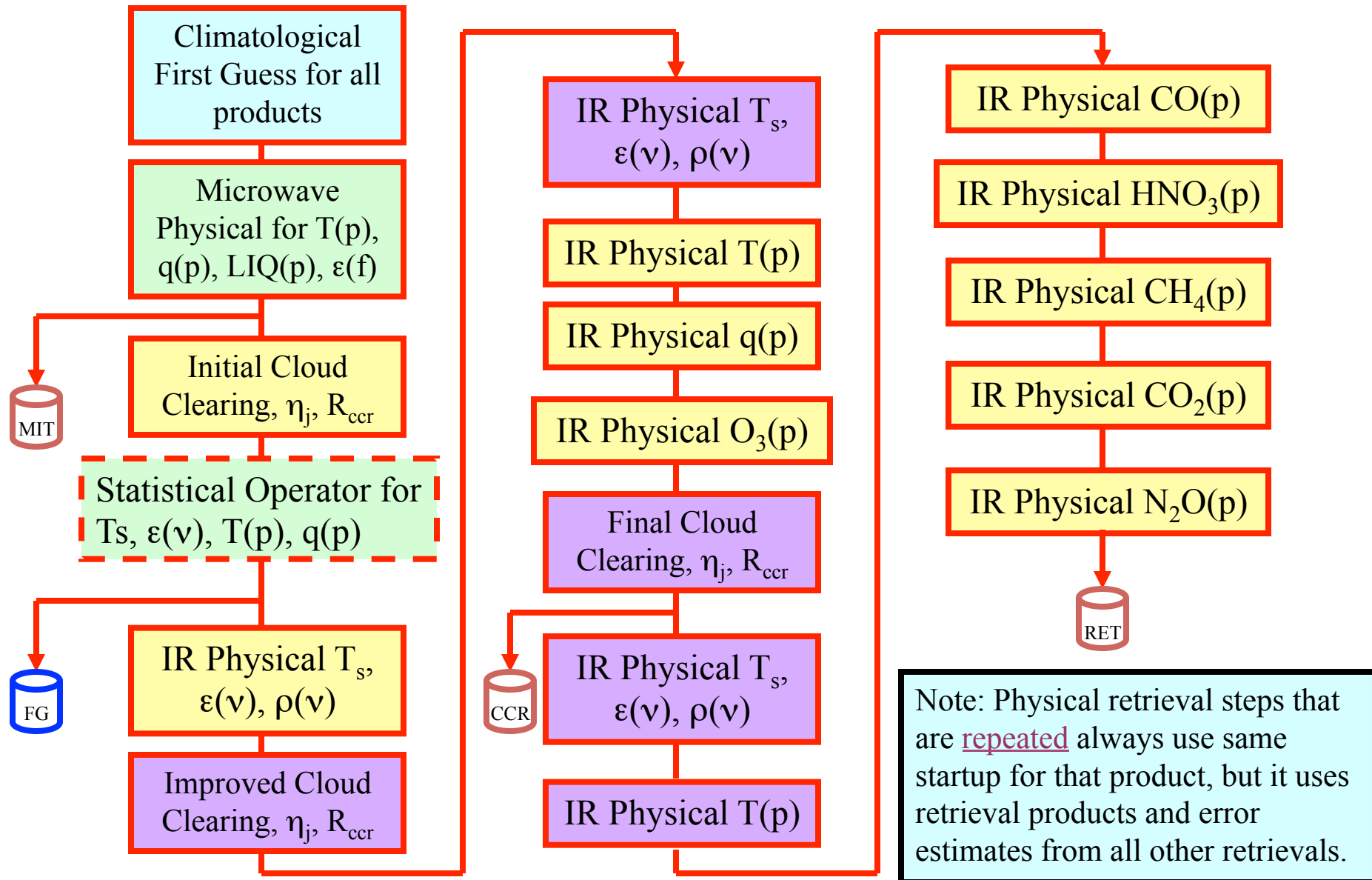
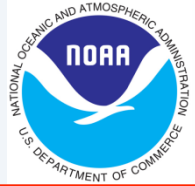
1DVAR versus AIRS Science Team Method

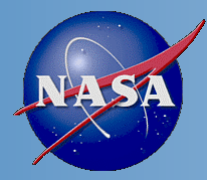


Simultaneous (1DVAR)	Sequential (AIRS method)
Solve all parameters simultaneously	Solve each state variable (e.g., $T(p)$), separately.
Error covariance includes only instrument model.	Error covariance is computed for all <i>relevant</i> state variables that are held fixed in a given step. Retrieval error covariance is propagated between steps.
Each parameter is derived from all channels used (e.g., can derive $T(p)$ from CO_2 , H_2O , O_3 , CO , ... lines).	Each parameter is derived from the best channels for that parameter (e.g., derive $T(p)$ from CO_2 lines, $q(p)$ from H_2O lines, etc.)
<i>A-priori</i> must be rather close to solution, since state variable interactions can de-stabilize the solution.	<i>A-priori</i> can be less complex for sequential with well selected channels.
Regularization must include <i>a-priori</i> statistics to allow mathematics to separate the variables and stabilize the solution.	Regularization can be reduced (smoothing terms) and does not require <i>a-priori</i> statistics for most geophysical regimes.
This method has large state matrices (all parameters) and covariance matrices (all channels used). Inversion of these large matrices is computationally expensive.	State matrices are small (largest is 25 $T(p)$ parameters) and covariance matrices of the channels subsets are quite small. Very fast algorithm. Encourages using more channels.
Has never been done simultaneously with clouds, emissivity(v), SW reflectivity, surface T , $T(p)$, $q(p)$, $O_3(p)$, $CO(p)$, $CH_4(p)$, $CO_2(p)$, $HNO_3(p)$, $N_2O(p)$	<i>In-situ</i> validation and satellite inter-comparisons indicate that this method is robust and stable.



Simplified Flow Diagram of the AIRS Science Team Algorithm

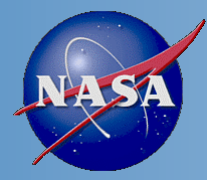




Advantages of the AIRS Approach



- Sequential physical algorithm allows for a robust and stable system with minimal prior information
 - Sequential approach allows the more linear parameters to be solved for first -- can make the algorithm very stable
 - Can solve for all significant signals in the AIRS radiances.
- Error from previous steps are mapped into an error estimate from interfering parameters
 - A unique feature of this algorithm is that error estimates from previous steps are mapped into subsequent steps
 - The observation covariance (S_ϵ in Rodgers 2000) contains both on- and off-diagonal terms composed of $(dR/dX) \cdot \delta x$ for all x 's that are considered interference (including cloud clearing, correlation due to apodization, etc.).
 - Can be more robust than simultaneous retrieval because each step uses optimal sampling of channels (*i.e.*, low interference).



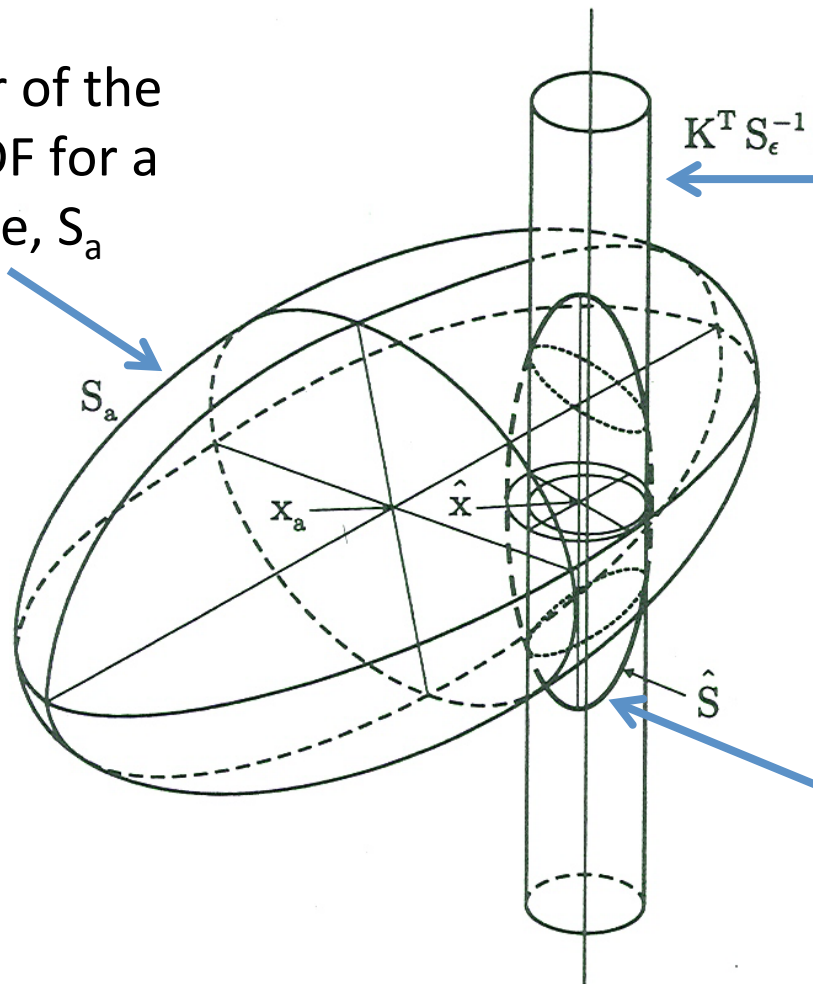
Advantages of optimal estimation



- O-E explicitly constrains the answer to lie within expectation of reasonable answers
 - Prior assumptions are always implicit in any retrieval approach
 - Note that “reasonable” can be in the eye of the beholder and sometimes that means a preference in the vertical null space.
- O-E explicitly derives the answer from prior information
 - in this sense, 1st guess can only speed up convergence
 - with enough iterations the same answer is usually achieved (up to non-linearity of Jacobians)
- Information content (or errors) in retrieval state can be partitioned between instrument and prior contributions
 - Averaging kernels or error covariance have more value

Graphical representation of O-E

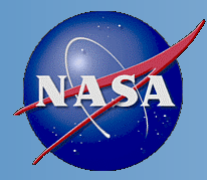
Contour of the prior PDF for a 3-D state, S_a



2-D measurement (i.e., no sensitivity to 3rd dimension) mapped into state space

Contour of the posterior PDF for a optimal retrieval.

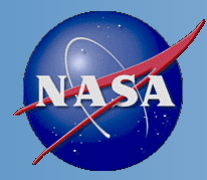
From (Rodgers 2000 World Scientific Publishing) Fig. 2.4 (pg.26)



Statistical Operators



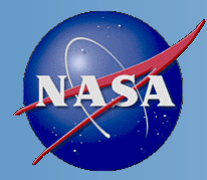
- Statistical retrievals are those that fit radiances, R , directly to an ensemble of geophysical parameters, X
 - $X = f(R)$, usually all radiances are used
 - Neural net: $X = A * \alpha(R) + B * \beta(R) + C * \gamma(R) + D * \delta(R)$
 - Linear regression: $X = A * R$
 - Neural Net has more free degrees of freedom
- Information can be derived from correlations
 - e.g. when we used to have an ozone regression we found that tropospheric ozone was being derived from AIRS channels sensitive to tropopause height and carbon monoxide
 - Would we call this a "measurement" or is it an "index"
 - We did learn from this – led to tropopause relative first guess



Training of AIRS statistical operators (global versus regional)



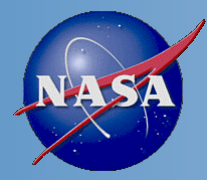
- NOAA regression was trained globally and used eigenvector regularization
 - We wanted to constrain the degrees of freedom allowed
 - 80 PCs with stratification into 4 view angle bins
- Neural net trained regionally, 200+ stratifications
 - 2 ascending and descending
 - 3 latitude bands (N.H., temperate, S.H.)
 - Each has frozen/non-frozen ocean, 5-7 surface pressure over land
 - 4 seasons
 - Version 6 Neural Net has *significantly* more free degrees of freedom to “fit” ECMWF
- Therefore, the differences between NOAA linear regression to MIT neural network approach can be do with these choices in stratification, constraints, etc.



Training of Statistical Operators (Geophysical Variance)



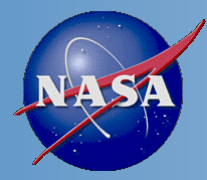
- Training must include every condition seen on Earth over the lifetime of the mission
 - For example, early in the AIRS mission we had issues with volcanic SO₂ from Etna
 - volcanic SO₂ was not in our early training (now it is)
 - Statistical operator extrapolated to completely unrealistic profiles
 - When it is good, it is very very good, but when it is bad
- Sub-resolved structure, being derived by correlations, needs expansive training
 - using ECMWF for training means we build in all ECMWF errors of the day
 - e.g., ECMWF ozone in May 2012 has very large errors
 - if this had been used for training of an ozone product it would have caused erroneous ozone products
- I would argue that there can never be enough training
 - Are there less obvious attributes of ECMWF that we have inadvertently embedded into our product?



Some concerns with the statistical operator have already been raised



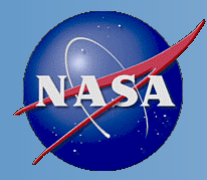
- Vertical structure has been shown to be greater than that which we can measure (Larrabee, Oct. 2011 AIRS meeting)
 - Statistical operator has ability to relate sub-resolved structure with AIRS radiances.
 - When the wrong structure is imposed in our first guess it is not removed by our physical retrieval (to be discussed in a few slides)
- Eric Maddy has shown that while Version 6 has significantly better statistics for temperature and water vapor profiles the cloud cleared radiance statistics are identical to v5.9
 - Implies that the improvement in $T(p)$ RMS may be due to sub-resolved vertical structures (*i.e.*, improvements in our null space, not our measurement)



Some mathematical issues with AIRS physical retrieval methodology



- We do not have a formal a-priori constraint.
 - We do have an ad-hoc "background term"
 - back in the day, I had convinced myself it provided the same functionality as a Roger's background term (recursively)
 - but this is not true, it does not equate to minimization of a cost function
 - iterations are done w.r.t. previous state, with some % held back
 - advantage: this retains the full vertical structure of first guess
 - disadvantage: there is no constraint, physical retrieval believes first guess
 - even if we characterized the statistical operator's covariance that information would not be used by our physical retrieval
- We only map the diagonal component of the error covariance into down-stream steps.
 - Eric Maddy has shown there is a robust way to pass the full covariance from one step to the next (Mar. 23 2007 AIRS meeting, my talk in session 6 and Eric Maddy Apr. 27, 2011 session 6)
- The physical algorithm has become a "QC" of the statistical operator
 - The goal is to select as many "good" cases and reject the "bad" regions
 - Usually, the statistical operator is very good (better than we can measure) so that "best" physical retrieval is one that does nothing
 - Tendency to over-regularize the physical retrieval



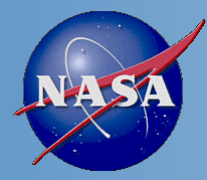
So, what is the most desirable system?



- If we fixed the “background term” then we must select a real prior state (need both state and covariance)
 - This can be non-trivial: for some products (or simultaneous “1DVAR-like” covariance) the covariance could be very difficult to construct.
 - Note that model priors also contain information on dynamics.

Prior information	Potential User Community
Statistical (with covariance)	Regional NWP
Climatology	Process studies
Forecast Model X (w/o AIRS R)	Global NWP for X, X=GFS,ECM,GMAO,etc.
Re-analysis product X (w/o AIRS R)	Historical climate for X

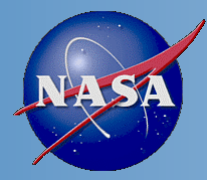
- O-E can also be done sequentially (and with cloud clearing) but for meaningful error estimates (or Averaging Kernels) we will need to improve the propagation of the error covariance downstream
- And there is a choice between clear-FOV retrievals (low daily yield, very good error characterization) or cloud clearing (high yield, complex error characteristics).



We could add more information content (i.e., minimize dependence on prior information)



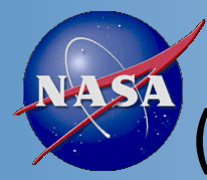
- MODIS radiances
 - NOAA already has MODIS IR convolved to AIRS FOVs
 - We also have AVHRR IR convolved to IASI (to be installed 2012)
 - ... and will have VIIRS IR convolved to CrIS (to be installed 2014)
 - It improves cloud clearing (part of our phase-2 IASI system)
 - Could potentially improve surface retrieval
- With degradation of AMSU and loss of HSB consider using alternative microwave radiances
 - CrIS/ATMS results demonstrate that the microwave information is important, especially for moisture
 - Eric has run ATMS+AIRS
 - Quick look results imply that the increase of information content may be more valuable than degradation of co-location
 - We could employ NOAA AMSU over life of AIRS mission



Validation



- "extraordinary claims require extraordinary evidence" Carl Sagan
 - We should avoid making algorithm choices using the same data sets used in “training” of algorithm or QC components.
 - We should partition improvements into those from null-space and those from physical measurement concepts
- Should the goal be to use IR everywhere?
 - Cloud clearing is known to fail in regions of high moisture or surface variability and has large non-Gaussian errors when it fails.
 - There is a trade-off between quality and robustness as scenes becomes more complicated.
 - CrIMSS metric is to have a retrieval everywhere
 - We look at both MW-only and IR+MW rets and decide where the IR retrievals have better performance.
 - To do this we must look at both accepted and rejected IR retrievals
 - We also require validation of a full profile (from TOA to surface).



Backup: O-E vs AIRS equations



(somewhat simplified to make them look similar)

O-E pivoting off of prior state:
$$X_j^i = X_j^A + \left[K_{j,n}^T \cdot N_{n,n}^{-1} \cdot K_{n,j} + C_{j,j}^{-1} \right]^{-1} \cdot K_{j,n}^T \cdot N_{n,n}^{-1} \cdot \left[R_n^{obs} - R_n(X^{i-1}) + K_{n,j} \cdot (X_j^{i-1} - X_j^A) \right]$$

Minimizes the cost function:
$$J = \left(R_n^{obs} - R_n(X_j^{i-1}) \right)^T \cdot N_{n,n}^{-1} \cdot \left(R_n^{obs} - R_n(X_j^{i-1}) \right) + \left(X_j^{i-1} - X_j^A \right)^T \cdot C_{j,j}^{-1} \cdot \left(X_j^{i-1} - X_j^A \right)$$

Equivalent to pivoting off of the previous iteration:
$$X_j^i = X_j^{i-1} + \left[K_{j,n}^T \cdot N_{n,n}^{-1} \cdot K_{n,j} + C_{j,j}^{-1} \right]^{-1} \cdot \left[K_{j,n}^T \cdot N_{n,n}^{-1} \cdot \left(R_n^{obs} - R_n(X^{i-1}) \right) - C_{j,j}^{-1} \cdot \left(X_j^{i-1} - X_j^A \right) \right]$$

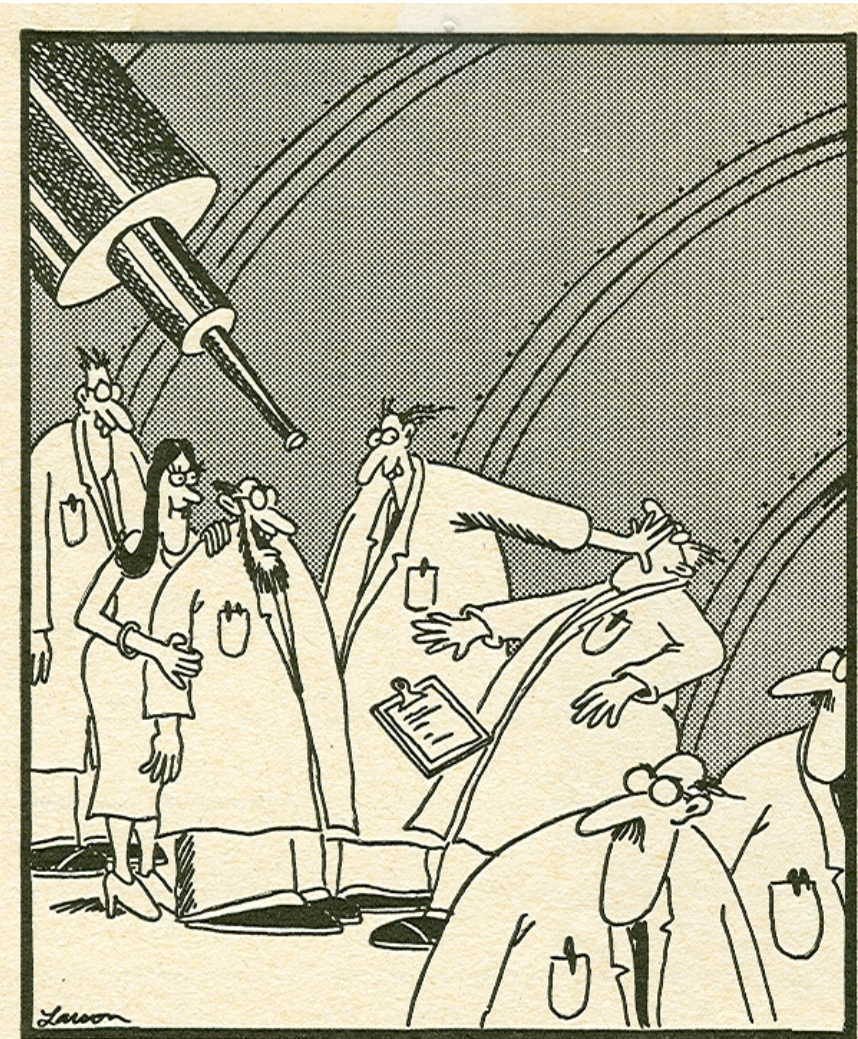
AIRS Science Team approach:
$$X_j^i = X_j^{i-1} + \left[K_{j,n}^T \cdot N_{n,n}^{-1} \cdot K_{n,j} + H_{j,j} \right]^{-1} \cdot K_{j,n}^T \cdot N_{n,n}^{-1} \cdot \left[R_n^{obs} - R_n(X^{i-1}) - \Psi_n^{i-1} \right]$$

H is a smoothing constraint and the background term is derived with respect to unregularized (LSQ) retrieval

$$\begin{aligned} \Psi_n^{i-1=0} &= 0 \\ \Psi_n^{i-1} &= K_{n,j}^{i-1} \cdot (X_j^{i-1}(0) - X_j^{i-1}) \end{aligned}$$

$$X_j^i(0) = X_j^{i-1} + \left[K_{j,n}^T \cdot N_{n,n}^{-1} \cdot K_{n,j} \right]^{-1} \cdot K_{j,n}^T \cdot N_{n,n}^{-1} \cdot \left[R_n^{obs} - R_n(X^{i-1}) \right]$$

- Suggested Rules for Engagement
 - suspend judgment
 - no speeches (1 minute rule)
 - one person speaks at a time (one idea at a time)
 - no killer phrases
 - hitchhiking is okay
 - be creative



All day long, a tough gang of astrophysicists would monopolize the telescope and intimidate the other researchers.